DOCUMENT RESUME

ED 037 793 CG 005 204

AUTHOR	Nitko, Anthony J.
TITLE	Some Considerations When Using A Domain-Referenced System of Achievement Tests in Instructional Situations.
INSTITUTION	Pittsburgh Univ., Pa. Learning Research and Development Center.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C.
PUB DATE	2 Mar 70
NOTE	24p.; Paper presented at American Educational Research Association Convention, Minneapolis, Minnesota, March 2-6, 1970
EDRS PRICE	EDRS Price MF-$0.25 HC-$1.30
DESCRIPTORS	*Achievement Tests, *Evaluation, *Learning Theories, Measurement, Measurement Techniques, Programed Instruction, *Testing Problems, *Test Interpretation

ABSTRACT

The problem of using a domain-referenced system of achievement tests is discussed as it relates to the design of instruction. Testing problems are discussed from the point of view that the teacher, pupil, and/or automation needs certain kinds of information in order to make instructional decisions that are adaptive to the individual learner. The design of achievement tests based on item forms is determined by the purpose(s) for which the information obtained from them is needed. The selection of items from the defined domain of item forms is discussed in terms of the purpose for testing, the relationship between items and instruction, and the relationship between instructional objectives and item forms. (author)

DOCUMENT RESUME

ED 037 793 CG 005 204

AUTHOR	Nitko, Anthony J.
TITLE	Some Considerations When Using A Domain-Referenced System of Achievement Tests in Instructional Situations.
INSTITUTION	Pittsburgh Univ., Pa. Learning Research and Development Center.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C.
PUB DATE	2 Mar 70
NOTE	24p.; Paper presented at American Educational Research Association Convention, Minneapolis, Minnesota, March 2-6, 1970
EDRS PRICE	EDRS Price MF-$0.25 HC-$1.30
DESCRIPTORS	*Achievement Tests, *Evaluation, *Learning Theories, Measurement, Measurement Techniques, Programed Instruction, *Testing Problems, *Test Interpretation

ABSTRACT

The problem of using a domain-referenced system of achievement tests is discussed as it relates to the design of instruction. Testing problems are discussed from the point of view that the teacher, pupil, and/or automation needs certain kinds of information in order to make instructional decisions that are adaptive to the individual learner. The design of achievement tests based on item forms is determined by the purpose(s) for which the information obtained from them is needed. The selection of items from the defined domain of item forms is discussed in terms of the purpose for testing, the relationship between items and instruction, and the relationship between instructional objectives and item forms. (author)

Some Considerations When Using a Domain-Referenced System of Achievement Tests in Instructional Situations

Anthony J. Nitko
Learning Research and Development Center
University of Pittsburgh

A paper presented at the Annual Meeting of the American Educational Research Association, Minneapolis, Minnesota, March 2-6, 1970.

The preparation of this paper was supported by the Learning Research and Development Center supported as a research and development center by funds from the United States Office of Education, Department of Health, Education, and Welfare.

Some Considerations When Using a Domain-Referenced System of Achievement Tests in Instructional Situations

In practice the use of item forms involves the employment of a sampling plan designed to generate test items that will ultimately appear on a test. A prior consideration to the sampling plan or item generation process, however, involves the question of the purpose for which a test is to be constructed.

It is difficult to separate the issue of "how to use item forms" from the issue of "for what to use item forms." When item forms are used to construct achievement tests that are to be used in different instructional systems, the students, teachers, content, and systems differ. It would seem, then, that each of these requires somewhat different kinds of achievement information and that each may require a different function to be served by the achievement information provided.

It is perhaps more meaningful then, to describe certain aspects of instruction and instructional design and to examine the possible use of item forms in these contexts, rather than to discuss the use of item forms in general.

The design of instruction may be considered to be centered around four general activities (Glaser, 1968): analysis of the subject-matter domain, diagnosis of the characteristics of the learner, design of the instructional environment, and evaluation of learning outcomes.

Analysis of Subject-Matter Domain

When analyzing the subject-matter domain for the purpose of instructional design, activities center around the specification of educational objectives and the translation of these objectives into some kind of assessable performance. At this stage, item forms serve the invaluable function of precisely defining the class of observations that will form the basis for planning instruction and inferring that the intended educational objectives have been attained by the learner.

Since item forms should define: (1) the instructions to the student, (2) the conditions for the performance, (3) the syntax and structure of the tasks, and (4) the manner in which the response is to be made, it is reasonable to expect that several item forms need to be constructed for each educational objective. Further, by specifying the parameters of the tasks a less ambiguous definition of an instructional objective is obtained. It is often recommended that sample test items accompany a verbal statement of an instructional objective. It would seem more useful, however, if each instructional objective were accompanied by several item forms which clearly specify the domain of tasks that are implied by the objective.

It should be pointed out that the item forms which have been reported to date (e.g., Hively, 1966; Osburn, 1968; Hively, Patterson, and Page, 1968) have more or less represented the content analysis of a given subject-matter area, rather than a behavioral analysis of it. If item forms analysis is to serve in the design of instruction, it must be formulated around the behavioral characteristics of the subject-matter. It is hypothesized that different item formats and different generation rules than are currently in existence would be forthcoming if behavior and content are taken into account.

Diagnosis of the Characteristics of the Learner

The second activity in the design of instruction consists of determining the relevant preinstructional characteristics of the learner for whom the instruction is designed. Some general classes of preinstructional variables have been specified by Travers (1963). The two we will consider here are: (1) the degree to which the terminal outcomes of instruction have been already attained by the learner and (2) the degree to which assumed prerequisites have been attained.

Often these two preinstructional variables can be designed into a single testing scheme. This type of adaptive testing is often called tailored or branch-testing. What is proposed here, however, is that the logic of the instruction and its prerequisites define the nature of the tailored test. This is a somewhat different use of tailored testing than has been reported in recent experiments (cf., Cleary, Linn, and Rock, 1968).

A recent pilot study by Ferguson (1969) is an example of how this procedure might work when it is coupled with item forms and a computer. (Time permits only a brief sketch of the procedure. Ferguson presented a complete description at an earlier session.) Table 1 illustrates terminal and prerequisite instructional objectives for an addition-subtraction unit from the elementary arithmetic curriculum of the Individually Prescribed Instruction Project (Lindvall and Bolvin, 1967). The unit is schematized in Figure 1. Each box represents one objective. The objectives are arranged in a branched hierarchy. Objectives 5, 17, and 18 are terminal objectives for the unit; the remaining objectives are prerequisites. Each of these prerequisite and terminal objectives was defined by one or more item forms which were then programmed for

use on the computer. The testing was done on an individual basis at a teletype terminal.

The object of the testing scheme was to locate a pupil at one of these objectives or "boxes" as quickly as possible and in such a way that he demonstrates mastery of objectives below his location and non-mastery of objectives above his location. The decisions for which the testing procedure must provide information are (1) what objectives should be tested and (2) does the pupil have mastery or non-mastery of the objectives that are tested. A decision needs to be made about every objective, but the trick is to make these decisions without testing every objective, and to minimize the testing for those objectives that are tested.

On this basis, a set of decision rules was devised that combined the capabilities of the computer with both statistical logic and subject-matter logic. This allowed "on-line" decisions to be made about what was to be tested and how extensively it was to be tested. The procedure breaks away from the traditional "test now, decide later" schemes that have received recent criticism (e.g., Green, 1969).

A decision about mastery of one objective that was tested was made by using the sequential probability ratio (Wald, 1947). An example of the situation is shown in Figure 2. The test length varies from pupil to pupil. A pupil is given only as many test items as are necessary to make a mastery or non-mastery decision with respect to a fixed mastery criterion and within prespecified Type I and Type II error rates. After each item is administered and scored, a decision is made to declare mastery, continue testing, or to declare non-mastery. With the number of items a randon variable, it is possible, in this example, to make a mastery decision with as few as 6 items and a non-mastery decision

with as few as 2 items. Not all mastery and non-mastery decisions are made this quickly; it depends on the response pattern of the pupil.

This figure illustrates the procedure for one objective. The problem that remains is that a decision needs to be made about every objective. Since the objectives are organized into a prerequisite sequence, the sequence itself can be used in the decision-making process. This results in the compound branching-rule shown in Table 2 for determining the next objective to be tested. The "next objective to be tested" depended on whether the student was declared a master or a non-master and on his response pattern that led to this decision. This is illustrated by the arrows sketched on the next figure (Figure 3).

Testing began at an objective in the middle of the hierarchy and continued until the branching-rule could not be satisfied. At that point, the object tested was the proper location of the student in the hierarchy. Untested skills could be assumed mastered or unmastered according to their position in the hierarchy and the student's response data.

An individual's testing session results in a profile similar to the one shown in Figure 4. The student would begin his instruction in this unit on the next sequential objective that was unmastered. In this example, he could begin either at Objective 8 or Objective 17.

Design of Instruction

The third task of instructional design is that of establishing procedures that allow the student to proceed from the "preinstructional state to a state of subject-matter competence" (Glaser, 1968). In his analysis of instructional design, Glaser specifies several conditions that influence subject-matter learning processes. These conditions

include: the sequence in which behaviors are learned, the stimulus and response properties of the learning tasks, the amount of practice and review that is necessary, the response contingencies related to error and correction, and the response contingencies related to effective reinforcement. Although it would be possible to elaborate on the use of item forms and item forms analysis in studying all of these conditions, we will discuss only response contingencies related to error and correction.

In some instructional designs, response contingencies related to learner error and correction are provided. If an instructional procedure is adaptive then learner errors are employed in the course of instruction. Since an item form defines a relatively homogeneous class of tasks, it would appear that the item form itself would be a useful diagnostic category in which error types could be studied. It is more likely, however, that an item form would need to be broken up into sub-item forms to yield relevant diagnostic information. This seems to be true particularly for those item forms that contain verbal material as variable elements--for example, the type of item form described by Osburn (1968).

Consider a similar item form presented in Table 3. This item form was used to generate a test that was administered to students in an elementary statistics course following a lecture on binomial probabilities. Before generating the test items, this item form was broken down into six strata. Each stratum contained a different verbal element from the set of elements defining the "region." The item formats defining each of these strata are shown in Table 4. The test contained 18 items in all; three items were randomly generated for each of the six strata.

Some simple results are presented in Table 5. It is seen that the various verbal replacement sets functioned differently. The items

from the first stratum were answered correctly by all students. The items from the other strata were nearly all of the same difficulty except for the items from Stratum III, Item 2 from Stratum IV, and Item 2 from Stratum V. When two items from Strata IV and V were more carefully examined it was seen that the particular numbers generated for the value of X_i appeared to influence the student's responses. The value of X_i generated for Item 2 of Stratum IV resulted in the region being defined as the entire sample space [$P(X \leq 4 | N=4, p=.20)$]. The value of X_i generated for Item 2 of Stratum V resulted in the region being defined as the null set [$P(X<0 | N=3, p=.25)$]. Thus, it seems that some of the numbers in the numerical replacement sets do not function "equivalently."

The errors that the students made along with their probable causes are shown in Table 6. Examination of student responses to the items from Stratum III indicated that those who erred primarily solved these problems as "equal to X" types or as "less than or equal X" types, rather than as a "greater than or equal to X" type. In this Stratum, Item 3 was answered correctly by slightly more students, but this was probably due to the particular number that was generated for X_i. The number resulted in $P(X \geq X_i)$ being equal to $P(X=X_i)$. It should be noted that on an individual basis, a student's erroneous responses within a stratum were quite consistent, thus allowing for a reasonably accurate error cause diagnosis.

Measuring Learning Outcomes

The fourth activity of instructional design involves defining means of measuring the outcomes of instruction. There are two general areas for which measures of learning outcomes can be constructed. One

set of outcomes consists of specified criterion behavior. These outcomes are assessed by examining student performance with respect to specified standards (Glaser, 1968). Such tests fall under the general category of criterion-referenced tests (Glaser, 1963). A second set of outcomes consists of specified constructs that are inferred from somewhat more broadly defined classes of behavior (Cronbach, 1969). Tests designed to measure construct outcomes generally fall under the category of norm-referenced measurement since all classes of tasks which define a construct can seldom be specified and, hence, interpretation of scores relative to norm groups have frequently been used. The validity of both of these kinds of tests have been considered in detail by Cronbach (1969) and Cronbach and Meehl (1955).

The use of item forms analysis is particularly relevant to the design of tests used to measure learning outcomes, regardless of whether the item forms are used to define the criterion behavior or as a basis for inferring the construct. The more precisely one defines the domain of test tasks the less ambiguous are the interpretations which are made from the results of testing. It seems important that item forms analysis be employed when criterion-referenced tests are used since absolute interpretations (cf. Cronbach, 1969) of test scores tend to be employed with this type of testing.

Item forms allow tests to be constructed along stratified sampling plans quite easily. The most obvious plan is to consider each item form as a stratum. Other stratifications should be considered--for example, stratification on the replacement sets within item forms. Comparisons of various families of stratified tests will indicate whether the equivalence of tests will be altered (Rajaratnam, Cronbach, and Gleser, 1965).

Such stratifications are also useful in redefining item forms if it is found that the within item forms strata substantially increase the generalizability coefficient.

If an existing test is subjected to item forms analysis a posteriori, the generic character (cf. Lord and Novick, 1968) of the test becomes apparent. Examination of the generic character of the test in terms of item forms would be useful in judging the adequacy of the test for the purpose at hand. Generic scores are useful in measuring instructional outcomes defined by constructs such as, reading comprehension, or arithmetic reasoning but they do not appear to be as useful for criterion-referenced tests or diagnostic tests of the type previously discussed.

Perhaps a word should be said about the elimination of items or item forms on the basis of statistical item selection procedures. Some questions were raised in this respect implying that elimination of items generated via item forms was not desirable (Osburn, 1968). It should be remembered that statistics are useful tools in the decision-making process, but do not replace the decision making itself. "Statistical considerations alone should not determine test design" (Rajaratnam, Cronbach, and Gleser, 1965, p. 54). Items generated by item forms can show "poor" statistical properties for many reasons including the way in which the item forms were written, the sampling plan used to select items and people for tryout, the content structure and the behavioral structure of the domain, and the purpose for which the test is used. As one example, consider a test designed to rank examinees with respect to their ability to recall basic addition and subtraction facts. It is easily seen how certain items (e.g., 1 + 1 = 2) could be eliminated from the final test designed to serve this ranking function. On the other hand, a criterion-referenced test designed to assess pupil performance with respect to

the domain of basic addition and subtraction facts would not employ statistical item selection techniques at all. One would want to assess the examinee on each and every basic fact during the course of instruction. Items would not be eliminated regardless of what the group statistics showed. If a test were to be designed to measure the construct, "arithmetic skills," in a broad population of children, one might well eliminate item forms dealing with basic facts entirely, since they would probably define the construct less well than other kinds of item forms and consequently allow individual differences to be reflected in the test scores.

Finally, it should be mentioned that item forms analysis allows the construction of tests that are designed to measure group parameters rather than individual parameters. Domains of tasks defined by item forms allow the matrix sampling techniques that have been advocated (see, e.g., Lord and Novick, 1968 for list of references; cf. Husek and Sirotnik, 1968) to be used to evaluate group outcomes. These procedures, which employ unmatched data, allow for many more observations on the domain than would otherwise be possible with matched data designs.

Concluding Remarks

This paper has attempted to discuss the problem of test design in the general context of instructional design. It has tried to show how item forms analysis can be used to construct test tasks that can be used to provide useful achievement data to the instructional designer as well as to the student and teacher. In short, item forms analysis should not be considered as simply a means of item generation, but as a procedure that allows the systematic study of the domain of instructional relevant tasks in terms of its structural and behavioral parameters.

The use to which the performance information will be put is the chief determiner of the way the item form will be written and the sampling plan used in constructing the test. It is probably true that for some types of measurement item forms are unnecessary to construct efficient tests. However, the questions that are always left begging when item forms are not employed are: "From what domain did these items come?" and "What are the behavioral characteristics of these items?"

Table 1

Terminal and Prerequisite Instructional Objectives
for an Addition-Subtraction Unit
in Elementary Arithmetic

(From Ferguson, R. L., 1969)

OBJECTIVE	BEHAVIOR
1	Solves addition problems from memory for sums $\leq$ 20.
2	Solves subtraction problems from memory for sums $<$ 9.
3	Solves subtraction problems from memory for two digit sums $\leq$ 20.
4	Solves addition problems related to single digit combinations by multiples of 10.
5	Solves subtraction problems related to single digit combinations by multiples of 10.
6*	Finds the missing addend for problems with three single digit addends.
7	Does column addition with no carrying. Two addends with three and four digit combinations.
8	Solves subtraction problems with no borrowing. Three and four digit combinations.
9	Finds the sum for column addition using three to five single digit addends.
10	Does column addition with no carrying. Three or four digit numbers with three to five addends.
11	Subtracts two digit numbers with borrowing from the ten's place.
12	Adds two digit numbers with carrying to the ten's or hundred's place. Two addends.
13	Adds two digit numbers with carrying to the ten's or hundred's place. Three or four addends.
14	Adds two digit numbers with carrying to the ten's and hundred's place. Two to four addends.
15	Subtracts three digit numbers with borrowing from the ten's or hundred's place.
16	Adds three digit numbers with carrying to the ten's or hundred's place. Two to four addends.
17*	Adds three digit numbers with carrying to the ten's and hundred's place. Two to four addends.
18*	Subtracts three digit numbers with borrowing from the ten's and hundred's place.

*Terminal objectives for this unit.

Table 2

Branching Rules for Computer-Assisted Placement Testing

Decision for 1 Skill	Pupil's Response Data (p)	Branching Rules (<u>Next</u> Skill to be Tested)
Mastery ($p \geq .85$)	HIGH ($p \geq .93$)	Branch <u>up</u> to <u>highest</u> untested skill.
	LOW ($.85 \leq p \leq .93$)	Branch <u>up</u> to skill <u>mid-way</u> between this skill and highest untested skill.
Non-Mastery ($p \leq .60$)	HIGH ($.43 \leq p \leq .60$)	Branch <u>down</u> to skill <u>midway</u> between this skill and lowest untested skill.
	LOW ($p \leq .43$)	Branch <u>down</u> to <u>lowest</u> untested skill.

Table 3

Example of an Item Form in Elementary Statistics which has a Verbal Replacement Set

Item Format

Let a person's score be the number of successes observed (N) independent trials of a Bernoulli experiment, where the probability of success on any one trial is (p). What is the probability that he would obtain a score (in Region B(N,p))?

Generation Rules

1. $N \varepsilon \{2 \leq N \leq 5 | N \text{ is an integer}\}$

2. $p \varepsilon \{.20, .25, .50, .75\}$

3. (in Region B(N,p))ε{of (X); greater than (X_i);

   at least as large as (X_i);

   less than or equal to (X);

   less than (X_i); between (X_i) and (X_j)}

4. $(X_i, X_j) \varepsilon \{0 \leq (X_i, X_j) \leq N\}; X_i < X_j$

   Note X_j is generated only when needed.

Sample Item

Let a person's score be the number of successes observed in 4 independent trials of a Bernoulli experiment, where the probability of success on any one trial is .75. What is the probability that he would obtain a score at least as large as 2?

Table 4

Sub-Item Form Strata Defined by the Various Verbal Elements of the Item Form Shown in Table 3

Stratum I.	Let a person's score be the number of successes observed in (N) independent trials of a Bernoulli experiment, where the probability of success on any one trial is (p). What is the probability that he would obtain a score of (X_i)?
Stratum II.	Let a person's score be the number of successes observed in (N) independent trials of a Bernoulli experiment, where the probability of success on any one trial is (p). What is the probability that he would obtain a score greater than (X_i)?
Stratum III.	Let a person's score be the number of successes observed in (N) independent trials of a Bernoulli experiment, where the probability of success on any one trial is (p). What is the probability that he would obtain a score at least as large as (X_i)?
Stratum IV.	Let a person's score be the number of successes observed in (N) independent trials of a Bernoulli experiment, where the probability of success on any one trial is (p). What is the probability that he would obtain a score less than or equal to (X_i)?
Stratum V.	Let a person's score be the number of successes observed in (N) independent trials of a Bernoulli experiment, where the probability of success on any one trial is (p). What is the probability that he would obtain a score less than (X_i)?
Stratum VI.	Let a person's score be the number of successes observed in (N) independent trials of a Bernoulli experiment, where the probability of success on any one trial is (p). What is the probability that he would obtain a score between (X_i) and (X_j) [inclusive]?

Table 5

Some Small Sample Data on the Stratified Item Form Shown in Tables 3 and 4

	Sub-Item Form Strata																		Total Test
	I			II			III			IV			V			VI			
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	
p	1.00	1.00	1.00	.91	.88	.88	.69	.53	.59	.84	.78	.91	.94	.79	.88	.84	.84	.84	
$\bar{p}$	1.00			.89			.60			.84			.85			.84			N=32
$\bar{X}$	3.00			2.66			1.81			2.53			2.56			2.53			15.09
s^2	0			.73			1.65			.67			.75			1.00			12.02

Total test: KR20 = .871*, Stratified KR20 = .933*

---

*<u>Note</u>: Test scores were skewed and sample size was small.

Table 6

Student Error Types for Various "Strata" Defined
by Sub-Item Forms Shown in Table 4

Stratum Type	Correct Solution Implied by Replacement Set	Item i: Student Error Type	Freq	Item j: Student Error Type	Freq	Item k: Student Error Type	Freq
II	$P(X>x_i \mid N,p)$	$P(X\le x^* \mid N,p)$	1	$P(X\ge x^* \mid N,p)$	1	$P(X\ge x_i \mid N,p)$	2
		$P(X=x_i \mid N,p)$	1	$P(X\ge x_i \mid N,p)$	1	$P(X>x_i \mid N^*,p)$	1
				$P(X=x_i \mid N,p)$	1	$P(X=x_i \mid N,p)$	1
				$P(X<x_i \mid N,p)$	1		
III	$P(X\ge x_i \mid N,p)$	$P(X\le x_i \mid N,p)$	7	$P(X\le x_i \mid N,p)$	10	$P(X\le x_i \mid N,p)$	9
		$P(X<x_i \mid N,p)$	1	$P(X=x_i \mid N,p)$	5	$P(X=x_i \mid N,p)$	2
		$P(X=x_i \mid N,p)$	1			$P(X=x_i \mid N,p^*)$	2
IV	$P(X\le x_i \mid N,p)$	$P(X\le x_i \mid N,p^*)$	1	$P(X<x_i \mid N,p)$	1	$P(X\le x_i \mid N^*,p)$	2
		$P(X<x_i \mid N,p)$	1	$P(X\ge x_i \mid N,p)$	4	$P(X>x^* \mid N,p)$	1
		$P(X=x_i \mid N,p)$	1				
V	$P(X<x_i \mid N,p)$	$P(X=x_i \mid N,p)$	1	$P(X\le x_i \mid N,p)$	5	$P(X\le x_i \mid N,p)$	2
		$P(X>x_i \mid N,p)$	1	$P(X\le x_i \mid N^*,p^*)$	1	$P(X<x^* \mid N,p)$	2
VI	$P(x_i\le X\le x_j \mid N,p)$	$P(x_i<X<x_j \mid N,p)$	1	$P(x_i<X<x_j \mid N,p)$	1	$P(x_i<X<x_j \mid N,p)$	1
		$P(X<x_i$ or $X>x_j \mid N,p)$	1	$P(x_i^*<X<x_j^* \mid N,p)$	1	$P(X>x_j \mid N,p)$	2
		$P(X=x_i \mid N,p)$	2	$P(X=x_i \mid N,p)$	1		

Note: 1. The symbols x*, p*, N* indicate that the student redefined the given values of x, p, N.
2. Computation errors (addition, subtraction, copying) are not shown.

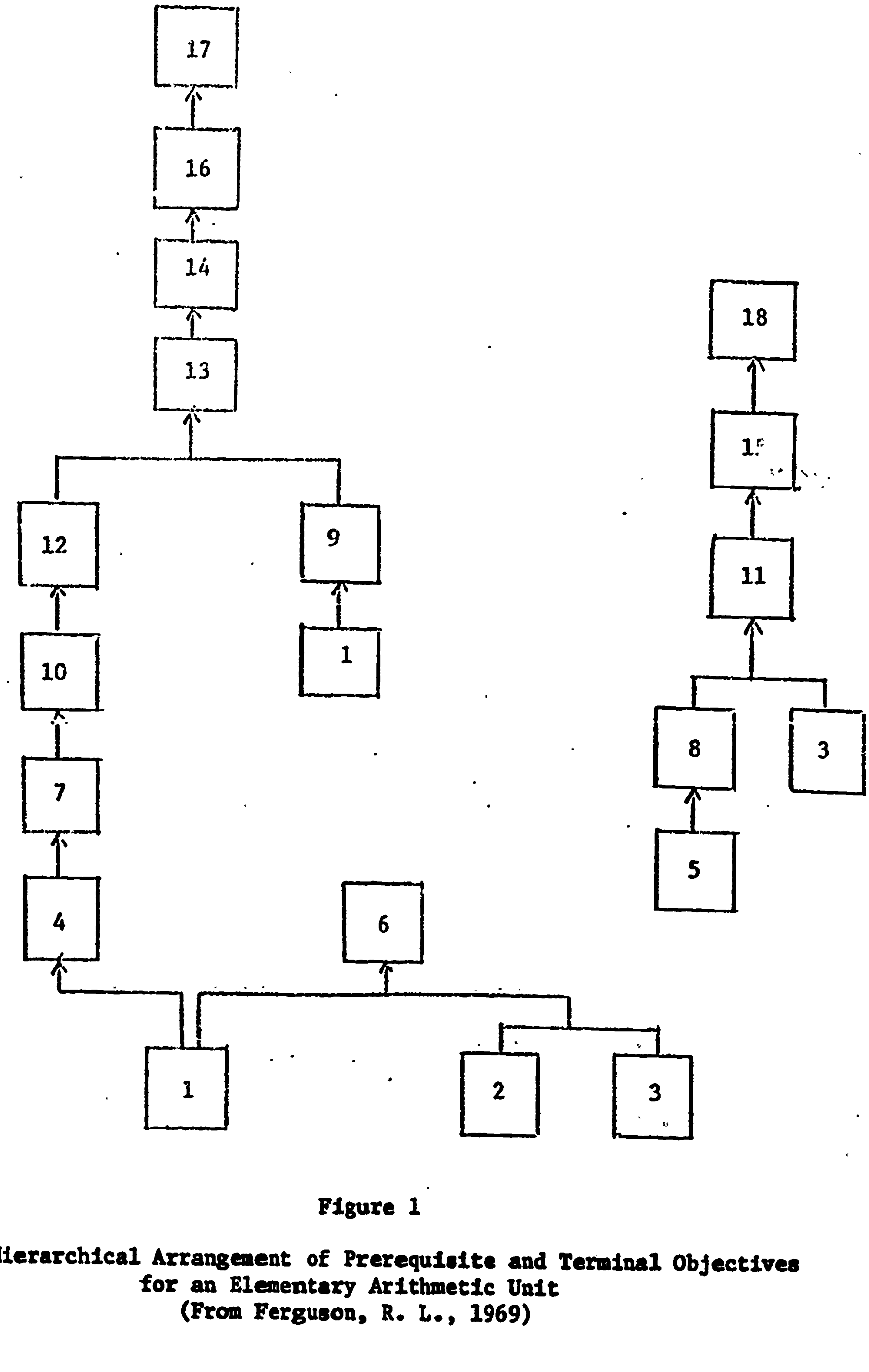
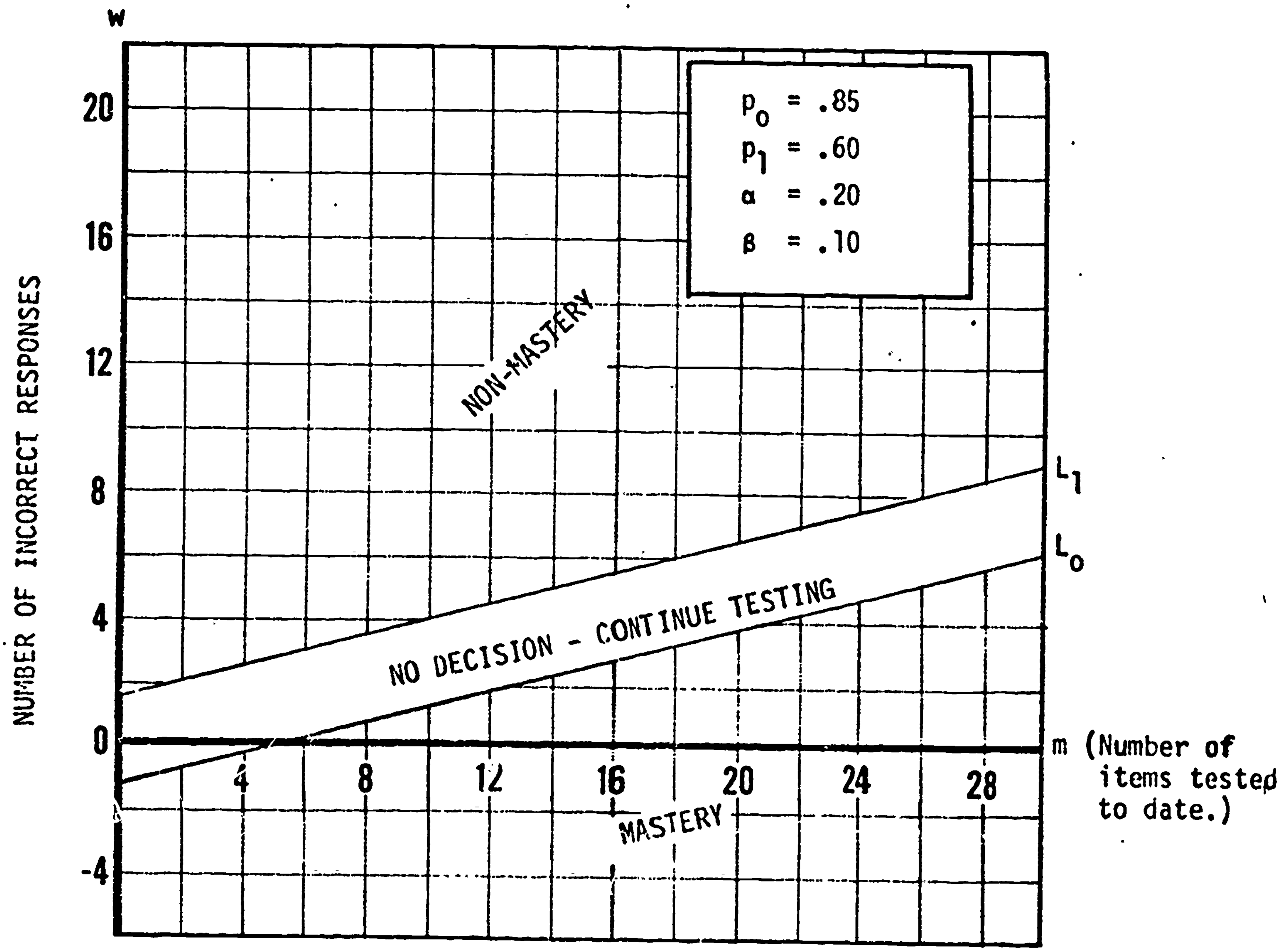


Figure 1

The Hierarchical Arrangement of Prerequisite and Terminal Objectives for an Elementary Arithmetic Unit (From Ferguson, R. L., 1969)



H_0: p=.85 (Student has sufficient mastery, omit instruction)

H_1: p=.60 (Student does not have sufficient mastery, give instruction)

Figure 2

Graph Illustrating Sequential Probability Ratio Test for Determining whether a Student does or does not need Instruction on an Objective (Modified from Ferguson, 1969)

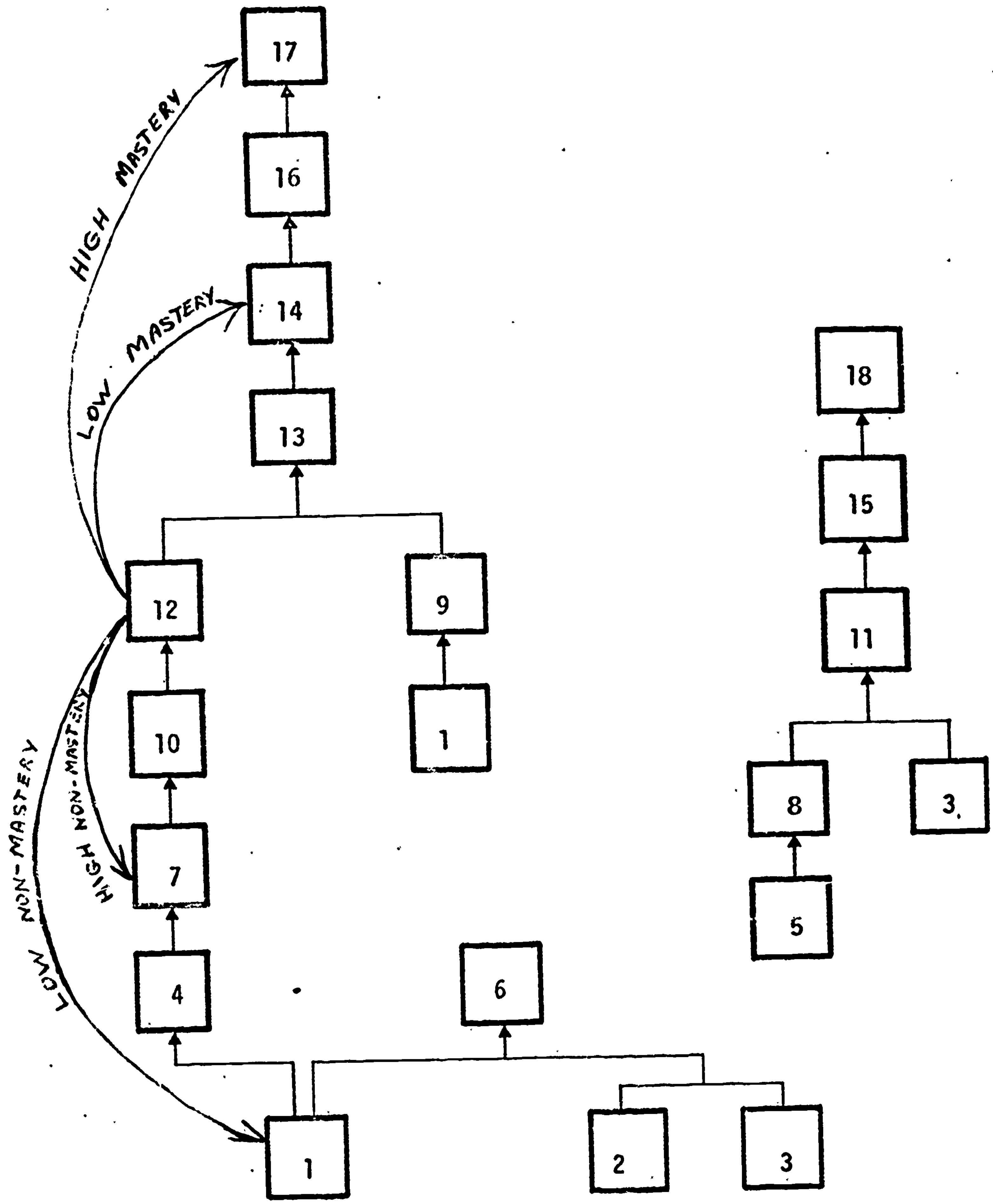


Figure 3

An Example of the Application of the Branching Rules of Table 2 to the IPI Mathematics Unit in One Instance

(Note only one of the "arrow" would be followed to locate the next objective to be tested. The branching rules would be reapplied after testing the next objective.)

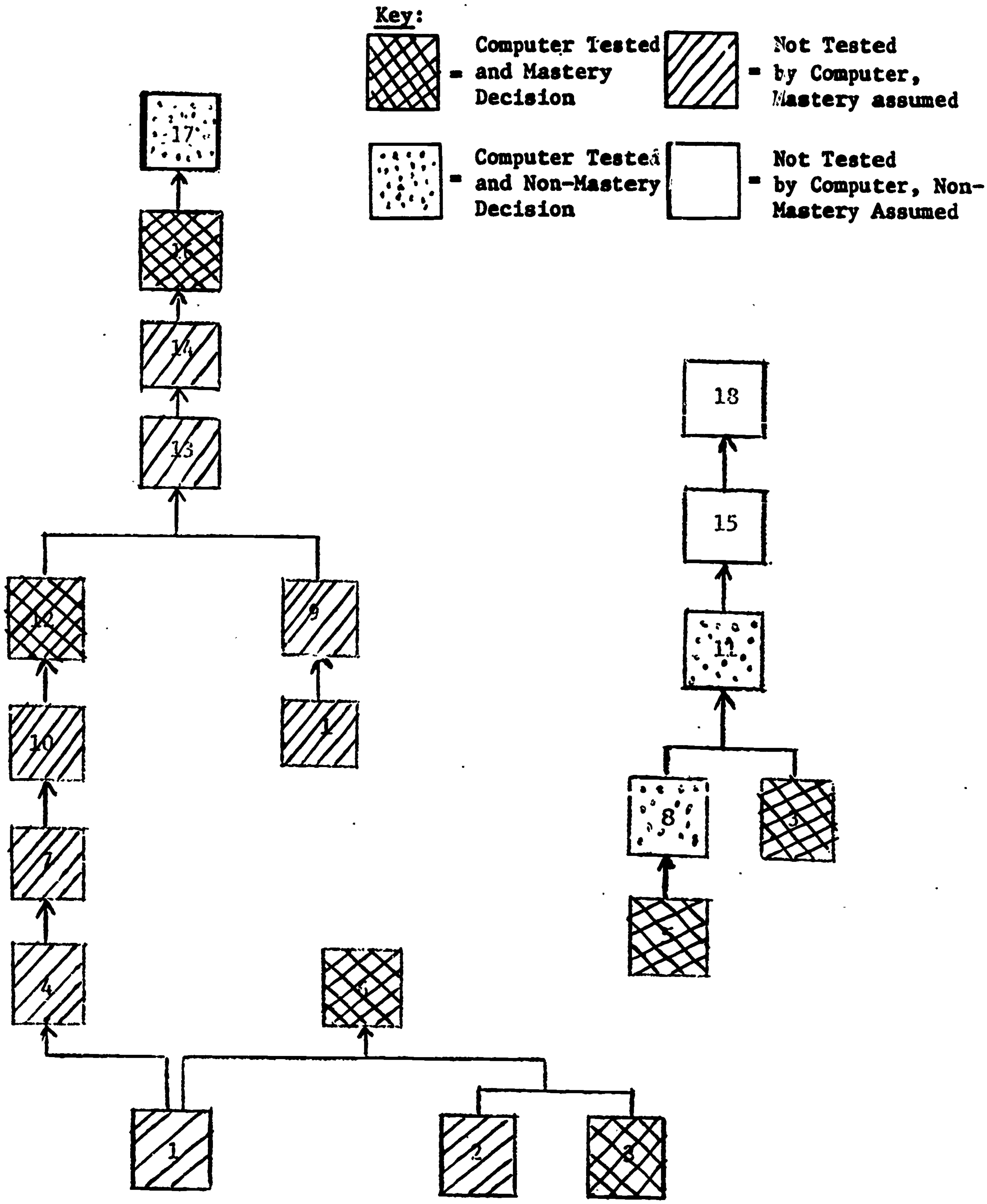


Figure 4

Example of a Student Profile Resulting from the Computer Testing of Prerequisite and Terminal Behaviors in an Addition-Subtraction Unit *

---

*The author is indebted to Dr. Richard L. Ferguson for making this data available to him.

References

Cleary, T. Anne, Linn, R. L., and Rock, D. A. An exploratory study of programmed tests. Educational and Psychological Measurement, 1968, 28, 347-349.

Cronbach, L. J. Validation of educational measures. Paper presented at the 1969 Educational Testing Service Invitational Conference on Testing Problems, New York, November, 1969.

Cronbach, L. J. and Meehl, P. E. Construct validity in psychological tests. Psychological Bulletin, 1955, 52, 281-302.

Ferguson, R. L. The development, implementation, and evaluation of a computer-assisted branched test for a program of individually prescribed instruction. Unpublished doctoral dissertation, University of Pittsburgh, 1969.

Glaser, R. Instructional technology and the measurement of learning outcomes. American Psychologist, 1963, 18, 519-521.

Glaser, R. The design and programming of instruction. The schools and the challenge of innovation, Supplementary Paper, No. 28. New York: Committee for Economic Development, 1968, 156-215.

Green, B. F. Comments on tailored testing. In W. Holtzman (Ed.), Computer-assisted instruction, testing, and guidance. New York: Harper and Row, 1969.

Hively, W. Preparation of a programmed course in algebra for secondary school teachers: A report to the National Science Foundation. Minnesota National Laboratory, Minnesota State Department of Education, 1966.

Hively, W., Patterson, H. L., and Page, S. A "universe-defined" system of arithmetic achievement tests. Journal of Educational Measurement, 1968, 5, 275-290.

Husek, T. R. and Sirotnik, K. Matrix sampling in educational research: An empirical investigation. Paper presented at the American Educational Research Association Convention, Chicago, February, 196 .

Lindvall, C. M. and Bolvin, J. O. Programed instruction in the schools: An application of programing principles in "Individually Prescribed Instruction." In P. Lange (Ed.), Programed instruction, 66th Yearbook, Part II. Chicago: National Society for the Study of Education, 1967, 217-254.

Lord, F. M. and Novick, M. R. Statistical theories of mental test scores. Reading, Massachusetts: Addison-Wesley, 1968.

Osburn, H. G. Item sampling for achievement testing. Educational and Psychological Measurement, 1968, 28, 95-104.

Rajaratnam, Nageswari, Cronbach, L. J., and Gleser, Goldine C. Generalizability of stratified-parallel tests. Psychometrika, 30, 1965, 39-56.

Travers, M. W. Essentials of learning: An overview for students of education. New York: Macmillan Co., 1963.

Wald, A. Sequential analysis. New York: Wiley, 1947.